Applied Artificial Intelligence
An International Journal

ISSN: 0883-9514 (Print) 1087-6545 (Online) Journal homepage: https://www.tandfonline.com/loi/uaai20

Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma

Martin Cunneen, Martin Mullins, Finbarr Murphy & Seán Gaines

To cite this article: Martin Cunneen, Martin Mullins, Finbarr Murphy & Seán Gaines (2019) Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma, Applied Artificial Intelligence, 33:3, 267-293, DOI: 10.1080/08839514.2018.1560124

To link to this article: https://doi.org/10.1080/08839514.2018.1560124

Published online: 25 Dec 2018.

APPLIED ARTIFICIAL INTELLIGENCE
2019, VOL. 33, NO. 3, 267–293
https://doi.org/10.1080/08839514.2018.1560124

Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma

Martin Cunneen[a], Martin Mullins[a], Finbarr Murphy[a], and Seán Gaines[b]

[a]Department of Accounting and Finance, University of Limerick, Limerick, Ireland; [b]Director of International Projects at Vicomtech-IK4 and Administrative Manager, Vi-DAS Consortia, San Sebastian, Spain

ABSTRACT

The question of the capacity of artificial intelligence to make moral decisions has been a key focus of investigation in robotics for decades. This question has now become pertinent to automated vehicle technologies, as a question of understanding the capacity of artificial driving intelligence to respond to unavoidable road traffic accidents. Artificial driving intelligence will make a calculated decision that could equate to deciding who lives and who dies. In calculating such important decisions, does the driving intelligence require moral intelligence and a capacity to make informed moral decisions? Artificial driving intelligence will be determined by at very least, state laws, driving codes, and codes of conduct relating to driving behaviour and safety. Does it also need to be informed by ethical theories, human values, and human rights frameworks? If so, how can this be achieved and how can we ensure there are no moral biases in the moral decision-making algorithms? The question of moral capacity is complex and has become the ethical focal point of this technology. Research has centred on applying Philippa Foot's famous trolley dilemma. We claim that before applications attempt to focus on moral theories, there is a necessary precedent to utilise the trolley dilemma as an ontological experiment. The trolley dilemma is succinct in identifying important ontological differences between human driving intelligence and artificial driving intelligence. In this paper, we argue that when the trolley dilemma is focused upon ontology, it has the potential to become an important elucidatory tool. It can act as a prism through which one can perceive different ontological aspects of driving intelligence and assess response decisions to unavoidable road traffic accidents. The identification of the ontological differences is integral to understanding the underlying variances that support human and artificial driving decisions. Ontologically differentiating between these two contexts allows for a more complete interrogation of the moral decision-making capacity of the artificial driving intelligence.

CONTACT Martin Cunneen martin.cunneen@ul.ie Department of Accounting and Finance, University of Limerick, Limerick, Ireland

All authors are members of the VI-DAS (Vision Inspired Driver Assistance Systems) H2020 research consortia.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uaai.

Introduction: The Problem Space

Driverless cars and automated vehicle technologies[1] (AVTs) have been headline news for some time. There is now a need to make a clear distinction when referring to the concept of driving. There is a need to identify, whether one is speaking of human driving intelligence (HDI[4]) or artificial driving intelligence (ADI[2]). If it is artificial intelligence controlling or driving the vehicle, how are we to understand its decisions, especially, the important decisions it makes in relation to its user's personal safety and the safety of others around it. Such decisions regarding safety and human welfare have an obvious moral weight but this does not mean that the ADI has moral intelligence or moral competence.[3] This pressing issue about machine morality and AVTs dominates recent literature, which has focused upon the question; how ADI responds to complex driving events that appear to require a moral decision. Is it even possible that ADI could have the moral intelligence required to make informed moral decisions?[4] If it can, is it important that we understand how? If it cannot, do users need to know why not, and recognise this as an identifiable limitation to the systems artificial intelligence and driving ability? This limitation could have a significant impact on the ADI's ability to make decisions in response to driving events such as unavoidable road traffic accidents[5] (URTA's) and influence public risk perception.[6] As Patrick Lin points out, ADI will most likely excel at driving decisions concerning manoeuvring from A to B. The improvements in manoeuvring and environmental awareness will constitute a passive phase of driving, focused on threat and collision avoidance. This passive phase of avoidance will constitute most of the ADI decision matrix. Nonetheless, Lin, Goodall and others point out the human road network will undoubtedly lead to scenarios where ADI is confronted with URTA's, that require a proactive decision matrix to respond to the complex scenarios involving decisions concerning loss of life (Gerdes and Thornton 2015; Goodall 2014a, 2016, Lin 2013, 2015, Trappl 2016). Accordingly, it is the decision capacity of ADI that is the focus of our analysis.

AVTs[7] present numerous decision challenges. Most hinge upon understanding the capacity and limitations of the AI. While the technologies emerge with ever increasing sophistication, it remains unclear how ADI will respond to complex unavoidable crash scenarios (Bonnefon, Shariff, and Rahwan 2016, Goodall 2014a, 2016, Lin 2013, 2015). The telos or goal of AVTs is to reach their destination safely and this will define AVTs as the technology continues to develop. The consideration of a driving telos[8] brings our attention to what many have identified to be a challenge that confronts the intelligence and decision capacity of the driving system. There is a need to identify the minimum level of ADI required, to achieve this telos of circumventing the human road network. One way of investigating this question, is to consider what accidentologist's refer to as driving events that interfere with normal driving phases (Hermitte 2012). Is it the case, that our human road network and environment are value loaded and will require an ADI to calculate human values? Children, cyclists, parents and babies, animals, and human actors are all part of the human road network and bring human values, rights and laws into the decision matrix of ADI. Numerous scenarios will confront ADI with what are complex value decisions and moral decisions. Current literature that focuses on the ethical analysis of moral decision-making capacities of ADI identifies that the decision-making capacity is far more limited than most are aware. It is often recognised as a logical rule-based system mainly determined by a judicial state's Highway Code (Dogan et al. 2016, Goodall 2014a). Therefore, does this mean that to achieve its telos, the ADI will demand greater artificial intelligence (AI) than one informed by object recognition and driving codes? Since its beginning in 1956 with John McCarthy,[9] AI was identified as the potential to artificially and digitally model human intelligence to carry out human tasks and workloads. We claim that the focus on AVTs and developing ADI presents another manifestation of the classical challenges that face AI.[10] The challenge concerns placing AI into human environments to replace human workloads with automation controlled by an AI. The challenge is evident when human decisions that are value centred, moral, or personal are intrinsic to a workload decision matrix, even in a secondary sense. This challenge has largely resulted in an inability to develop an AI to cater for the complex array of human intelligence and human decision-making processes. This same classic AI scenario now confronts ADI with URTA's, as an example of a human driving workload, presenting complex moral challenges. These confront ADI with decisions that require more than intelligence concerning manoeuvring the vehicle can support. For example, how can human morals, laws, systems of value and distinct human ethical theories that are meaning specific to human intelligence and human social contexts, be applied to ADI[11]?

From considering accurate risk metrics, ethical analysis,[12] private, and public policy on conduct and behaviour, responding to public risk[13] and safety perception, the concept of AVTs decision-making systems will be integral to defining the technology. The problem is that the decision spectrum and, specifically, the moral decision capacity remain unclear. This is most evident in considering the emotive scenario of URTA's. There is a need to clarify what the emerging possibilities of ADI are in relation to its decision capacities and limitations. We claim that the decision matrix of HDI and ADI constitute two distinct paradigms of intelligence that determine two separate ontologies of decision-making in relation to URTA's. Each paradigm, presents a spectrum of intelligence that signifies the capacity of the intelligence to make decisions. This is especially evident when one considers HDI as incorporating human moral intelligence. HDI and ADI, in relation to the capacity to make moral driving decisions are intrinsically different. To apply human moral experiments, theories and reasoning to ADI without this awareness will lead to misconception. To develop informed moral frameworks of analysis that support the development of ADI, we argue that the two paradigms of intelligence and the ontologies of decision-making must precede ethical analysis. It is only by first considering the differences in the ontologies of HDI and ADI that one can perceive how possible different manifestations of ADI and the addition of artificial moral intelligence, could also further determine the ontology of the scenario. This complex spectrum of ontological differences will inform all conceptual frameworks of meaning. Key concepts such as accountability, liability, risk, intentionality, agency, and responsibility will all be determined by a context of meaning that is specific to the relative ontology of the scenario.

The paper consists of two stages. The first follows the introduction in setting out the complex relations of the problem space. It identifies the existent literature attempts to apply human ethical theories and human moral intelligence to A.I as an example of ontologically deficient analysis of AVTs and ADI decisions and specifically moral decisions in response to URTAs. The second part defends the claim that ADI decisions present a unique ontological input to ADI actions. To support this, the analysis focuses on mapping driving decision matrixes of HDI and ADI. It then identifies the differences in how HDI and ADI as paradigmatic differences. The identifiable differences in the conceptions of intelligence point to the impossibility of ADI having moral intelligence in the same form as human intelligence has moral intelligence.

Artificial Driving Intelligence and Trolley Dilemma's

The article responds in part to what we claim is a lack of consideration of decision ontology in investigating machine decisions and the question of moral decision capacity. So much so that the current literature focus on ADI and artificial moral intelligence has neglected to consider the problem space in terms of ontology. The most apparent reason for this failure relates to the understandable human context of ADI. There are so many humanising aspects to ADI, it is difficult to appreciate that the differences are so great that once identified they outweigh the similarities. ADI is coloured by human design, it caters for human workloads, and tasks are enveloped in a distinctly human road network and environment. Therefore, not surprisingly, there is a consistent effort to apply human decision and moral decision models of analysis to artificial intelligence decision metrics and the possibility of artificial moral decision intelligence. The use of traffic models to elucidate accident scenarios is generally accepted as a fruitful means of elucidating the relations of agents in an accident scenario. However, using trolley dilemmas to portray URTA's in the context of ADI decisions and moral decisions is not a straightforward application. We claim there is not an isomorphic relationship between HDI and ADI that support's the application of the trolley dilemma to ADI. The important difference concerns the differences in the ontological form of relations between each intelligence and the objects it perceives, identifies, and makes decisions in relation to such assessments. There remains a need to investigate and frame the decision-making capacity of ADI as an intelligence that determines a unique accident and driving ontology. This can only be achieved by comprehending the difference of its workings by comparing them to HDI decisions. Human and artificial driving decisions are intrinsically related by sharing the same telos of safe efficient driving on our human road network. In this sense, the goal and the environment are both the same. They are also similar in how each relies on the sensory visual perception of the environment to obtain information regarding the terrain and the populated objects in it.[14] This information is assessed and processed by intelligence that evaluates plans and forwards actionable outputs as driving decisions.[15] To some extent, ADI is necessarily an anthropomorphic model of HDI given that vehicle design, the road network and driving code of conduct (Highway Code), legislation and supporting features, such as insurance and driving ability, are all context specific to the human phenomenon of driving. The surface similarities here can be deceptive given that the information and the processing regarding each are categorically different. The similarities are just that, nothing more than similarities that can at a superficial level, offer assistance in how we conceptualise ADI decisions. However, beyond the superficial, the categorical differences are we argue, equal to two paradigms of intelligence that ontologically determine decision-making matrixes. HDI is the paradigm of organic intelligence that reflects a human intelligence applied to a particular task or workload of driving. Whereas ADI is the paradigm of AI for driving, it carries out specific technological envelopes of operation that collectively constitutes what is a topographical representation of the phenomena of driving. The surface similarities can be captivating and can distract from the significance of the differences between the two paradigms of driving intelligence.[16]

The trolley type dilemmas are thought experiments designed to communicate the difficulties of moral conflicts between core ethical beliefs, our social and individual moral spectrums that point to moral relativism, the conflicts between important cornerstone values such as its right to save the greatest number and its right to support the principle that killing is wrong.

Thomson developed the analysis and popularity of trolley type problems further (Thomson 1985). Her account is perhaps the most succinct:

> *"Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit of a valley at that point, and the sides are steep, so you must stop the trolley if you are to avoid running the five men down. You step on the brakes, but alas they don't work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the straight track ahead. Unfortunately, Mrs. Foot has arranged that there is one track workman on that spur of track. He can no more get off the track in time than the five can, so you will kill him if you turn the trolley onto him. Is it morally permissible for you to turn the trolley?"* (Thomson 1985)

Thomson maintains that all the parties she asked would turn the trolley to save the five and kill the one workman (Thomson 1985). She identifies the conflict between the core ethical belief it is wrong to kill, and it is wrong not to save lives if you can. She appeals to a contrasting "transplant" scenario of a surgeon who can save five lives at the expense of one life as an organ donor. She contends that the surgeon is confronted with the same emotive scenario to save five at the expense of one life. Thomson appeals to Foot's (1978) analyses and elucidates the important difference between the trolley driver and the surgeon, as the former's inaction causes five deaths, whereas, the surgeon's inaction is not the cause of the five deaths (Thomson 1985). It becomes apparent that the subtle differences between what appear as quantifiable scenarios, are important in determining the moral weightings of the scenario. The intent of the actor determines the moral weight of the scenario. It shifts from the surgeon seeking to save more lives at the expense of one, to the surgeon attempting to rectify his previous error. The intent is key in determining the contextual meaning of the scenario. The thought experiment brings to the fore, the contrast between duty ethics (deontological ethics) such as the Kantian universal law (Kant 1783), we should not kill, contrasted with the general principle of utility (Utilitarian, Consequentialist) approach, it is the greatest happiness (and absence of pain) for the greater number that makes it permissible to kill one to save five (Bentham 1843).

Thomson identifies how the trolley dilemma proved useful in differentiating between two quantifiably similar scenarios wherein one person is positioned against five with the same outcomes, one person's life forfeited to save five. Thomson identifies there are numerous possible and subtle differences that can dramatically affect the ethical context of the same quantifiable scenario. For example, if the transplant scenario is further weighted in terms of the surgeon correcting previous errors he had made that led to the five people requiring lifesaving transplants then the scenario is weighted differently. Thomson identifies there are always variable extenuating factors that can dramatically change the contextual weighting of the dilemma in how we perceive the ethical meanings. When considering Thomson's insight into the trolley dilemma in relation to AVTs, the differences the driving intelligence brings to the URTA's will have a significant effect upon how we interpret and understand the moral context and meaning of the scenario. What the different human drivers bring to an URTA determine the ontology of the scenario, for example, one human driver would value not hitting the fox or deer on the road while another will have no problem hitting it. The action of each reflects their moral intelligence, not that one is more morally astute than the other but rather as a moral intelligence that will or will not value the animal as an object to be killed or not. What this communicates concerns object value identification, in the topographical map, this constitutes one aspect of the information but the important relations between the objects constitute another. For driving intelligence to achieve its driving goal it, therefore, requires the second aspect of information relating to object dynamics and relations. The trolley dilemma, when applied to HDI, is useful in elucidating the subtle differences in what agents bring to a scenario. How agents represent and interpret a scenario. It also elucidates emotive moral values such as how we judge a morally reprehensible and difficult decision. However, the utility of the trolley dilemma is not strictly transferable to ADI. Nonetheless, its popularity in engineering and programming means that it is most likely to continue as a thought experiment. Therefore, it needs at very least to be first informed by some ontological elucidation of the differences brought about by different types, degrees and forms of intelligence. The trolley dilemma has been used by numerous ethicists and technologists to elucidate the conceptual framework of ADI confronted with URTA's. A common example of the trolley dilemma applied to ADI is forwarded by Lin below. Lin's application communicates the confusion as one of seeking to address HDI and its unique contextual meaning to be isomorphic to ADI.

> *"Your car detects a pickup truck coming up behind you, about to cause a rear-end collision with you. The crash would likely damage your car to some degree and perhaps cause minor injury to you, such as whiplash, but certainly not death. To avoid this harm, your car is programmed to dash out of the way, if it can do so safely. In this case, your car can easily turn right at the intersection and avoid the rear-end collision. It follows this programming, but in doing so, it clears a path for the truck to continue through the intersection, killing a couple children and seriously injuring others."* (Lin 2015, 78)

Patrick Lin (2013), following Wallach and Allen (2008) and Bryant Walker-Smith (2012), appeals to the trolley dilemma to argue that machine morality will be required for a "robot car" to adequately respond to driving events and specifically URTA's (Lin 2013, 2015). Lin appeals to numerous hypothetical scenarios (above) to communicate and defend his claim; the human road network will confront ADI's with scenarios that would benefit from artificial moral intelligence. From Lin's research, the field of AVTs ethics has grown;

several other commentators have emerged in further developing the area. Noah Goodall confronts the question of AVTs and URTA's, while also attempting to elucidate the technologies that could support machine morality and their potential limitations (Goodall 2014a).

> *"If, however, injury cannot be avoided, the automated vehicle must decide how best to crash. This decision quickly becomes a moral one…"* (Goodall 2014a, 60)

In his latest work, Goodall (2016) highlights what has become a common concern regarding the negative media attention surrounding AVTs and URTA's as trolley dilemmas. Goodall points out that the trolley dilemma has a positive utility in two important ways; first, it offers an elucidatory method that assists in identifying and exploring our moral reasoning and intuition. Second, it offers a means of developing what Goodall describes as "edge cases" that are important in forwarding systems of analysis by contrasting hypothetical scenarios (Goodall 2016). The trolley dilemma has also received criticism as mere armchair philosophising that is remote from the reality of the technology (Goodall 2016, Nyholm and Smids 2016, Rose 2016). As Goodall acknowledges, there is now a general disdain toward the trolley dilemma. This disdain is also countered by Lin who maintains that no matter how unrealistic the trolley dilemma appears it remains a problem that programmers of ADI must confront:

> *"If you complain here that robot cars would probably never be in the Trolley scenario – that the odds of having to make such a decision are minuscule and not worth discussing – then you're missing the point. Programmers still will need to instruct an automated car on how to act for the entire range of foreseeable scenarios, as well as lay down guiding principles for unforeseen scenarios."* (Lin 2013)

The trolley dilemma highlights this difficulty concerning pre-crash algorithms and decision making, regarding human life. Applying the trolley dilemma to ADI decisions regarding the URTA and assessing impact strategies[17] to lessen life loss or injury, is more difficult to comprehend. The moral experiments that trolley dilemma's present have been favoured by the media, shunned by AVT industry, and the focus of debate by robot ethicists (Lin 2013, Wallach and Allen 2008), psychologists (Bonnefon et al. 2015, Bonnefon, Shariff, and Rahwan 2016), and technologists. As Lin (2013) points out, at very least it offers insight into the subtle differences between our moral intuitions and moral theories. However, the trolley dilemma can lead to misaligned strategies of ethical analysis. This is primarily due to the dogmatic applications that take the surface similarities to be definitive without cognizance of the important differences between HDI and ADI. This is the focus of the analysis. The trolley dilemma does not define the ethical space in relation to AVTs, but it offers a platform for elucidating the important differences between HDI and ADI. This point is developed by

Goodall who places more focus on the assessment of Risk as a more conducive platform for analysis.

> *"While the trolley problem is valuable in isolating people's intuitions about morally ambiguous crash decisions and stress testing ethical strategies, it represents a fairly narrow area of automated vehicle ethics and suffers from a perceived lack of realism."* (Goodall 2016)

In one context, the trolley dilemma, in relation to AVTs, present what Bonnefon et al. (2015) describe as the "flagship dilemmas of experimental ethics". The trolley dilemma also has important applications that move beyond the ethical context, to questions relating to frameworks of governance in relation to legal issues, accountability, responsibility, liability, risk perception, regulation and risk analysis (Coeckelbergh 2016, Hevelke and Nida-Rümelin 2015). The trolley problem is either a beneficial ethical exercise that elucidates the genuine limitations or problems that AVTs present (Bonnefon et al. 2015, Goodall 2014a, Lin 2015) or it is a captivating ethical experiment with little value to experimental ethics and the analysis of AVTs, (Nyholm and Smids 2016). While Nyholm and Smids are more critical of its application and argue there are important disanalogies between the trolley dilemma and its application to AVTs.

> *"We think, therefore, that it is important to resist the temptation to draw a very strong analogy between the ethics of accident algorithms for self-driving cars and the philosophy of the trolley problem."* (Nyholm and Smids 2016)

In contrast to the above, we claim that the trolley dilemma has a dual role. First, it will act as an ontological experiment, supporting the assessment of hypothetical scenarios that permit explication of the subtle relationships between objects, value identification, and the intelligence assessing them. Second, it will inform the ontological map it can have a utility in revealing moral perspectives, conflicts, and expectations. It will not define the landscape, it will elucidate it. For the trolley dilemma to support accurate analyses of ADI, it must be informed by the ontological differences between HDI and ADI. There are several ontological points of interest that are intrinsic in framing the relationship between intelligence, object awareness, identification, and decision processing. We claim these points are integral to understanding the technology and provide an accurate informed basis upon which further conceptual schemes of analysis must be guided by. Therefore, without considering the ontological differences and their import upon conceptual frameworks, subsequent analysis of the technology, especially ethical analysis, will run the risk of misapplication. The analysis utilises the trolley dilemma, not as a moral experiment per se but rather as a hypothetical tool to elucidate ontological differences between HDI and ADI intelligence and decision matrixes. Therefore, before anticipatory ethical analysis can take place, it must be informed by a specific ontological analysis of how AI determines the ontology of the scenario.

Intelligence, Object Identification, and Ontology

One of the key differences between HDI and ADI resides in what these two entities bring forth to the accident scenario. What HDI and ADI bring, in capacity and abilities to make decisions[18] in response to accident scenarios, will in part relate to what values are placed upon the entities populating the scenarios (Dogan et al. 2016). A key component of the emerging AVTs concerns the ability to visually identify and relate objects on a topographical map (Chen et al. 2015). The ADI must have the capacity to perceive its "*natural environment*" and make "*intelligent decisions*" (Cheng 2011). It is the capacity to identify objects and the ability to process input data to support actionable outputs that define ADI. The decision-making process "*consists of mission planning and behavioural reasoning*" which will need to be updated in real-time to account for the changing environment (Cheng 2011). Four key component technologies define an AVT or "*intelligent vehicle*" these are: "*environmental perception and modelling, localisation and map building, path planning and decision making, and motion control*" (Cheng 2011). The intelligence of an artificial driving system is composed of the symbiotic relationship of the four technologies. The ability of ADI to identify objects and ascribe possible value data to the objects represents different levels of intelligence of a system[24] (Chen et al. 2015). The levels of ADI relate to the functionality of the system and the decision metrics that determine its responses and actions (Goodall 2016). System sophistication depends on the data the system receives (input) and the capacity of the system to process the data into information supporting a decision and action (output). For example, if an ADI can only identify objects as material bodies with classifiers relating to size and movement, this capacity limits the value decision metrics to manoeuvrability, avoidance, and potential risk identifiers relating to collisions. This level of ADI utilises the topographical map, as an augmented reality consisting solely of objects, there is no application of classifiers to distinguish different organic beings or to calculate and place value and risk metrics on to them. With the addition of values that relate to human values of life, the level of intelligence required is more complex than classifying objects according to shape, size, and motion that are informed by a state's Highway Code. Every addition of value data to objects will increase the complexity of the classifiers required to support the capacity of the decision-making intelligence to respond to events. Increasing availability of object data increases the capacity of the ADI to make decisions that are more complex. It also increases the burden of responsibility and risk of culpability in forwarding pre-collision decisions. Therefore, the instantiation of intelligence represents a point on a spectrum of ADI that ranges from the above object recognition to object recognition with numerous additional dynamic values. The object recognition capacity is an integral part of the intelligence system that is intrinsic to the capacity to make decisions. There is a need to investigate instantiations of ADI as the technology develops. This presents an ongoing obligation to ensure ADI decision capacity can sufficiently respond to the complex array of driving events it will encounter in traversing our human driving environments. The challenge will relate to the ability of a system or intelligence to classify objects in an environment. Object recognition and complex decision making first presents a classification problem and then an intelligence problem. Classifiers will be the first development of the technology to deliver the required data that the system can then analyse to determine decision responses.

Decision metrics depend on object identification and classification; these will in part determine the effectiveness of the technology to respond to complex scenarios or driving events. As highlighted in the chart, the increasing addition of values will point the development of the technology to an ethical outcome (Dogan et al. 2016). We argue that in the context of driving event decisions, especially concerning unavoidable road traffic accidents URTAs, there is an underlying ontological scale of meaning that is intrinsic to the conceptual framework of meaning that defines our understanding of an event. Elucidating the ontological scale is integral to developing accurate conceptual meanings, frameworks, and specific conceptual applications. For example, investigations into driving events that relate to questions pertaining to liability, intentionality, accountability, and responsibility, need to address the ontological status of the event in question. To assess the intentionality of decisions, the intelligence and the workings of the decisions must be evaluated. Decision intelligence and decision intentionality are key factors that determine the ontology of the scenario. Others also emphasise the importance of ontology and object identification (Bonnefon, Shariff, and Rahwan 2016), ontology and AI (McCarthy 1996) and ontology in relation to accidentology (Hermitte 2012, Van Den Beukel and Van Der Voort 2016, Wang and Wang 2011). The paper defends a more thorough investigation of the significance of ontology to AVTs by contextualising ontology in relation to AI decisions, HDI and ADI decisions, and specifically, ADI pre-collision driving decisions. We argue that to acquire informed and accurate knowledge relating to ADI decisions and driving events, it is necessary to identify the specific ontological status of the decisions. To achieve this, it is necessary to approach ADI from the standpoint of two important components we claim frame the ontology of driving decisions. The first concerns the objective ontology, it relates to what entities populate the environment and how the actors, objects, and artefacts determine the ontology. For example, the ontology of events in inner city rush hour driving is dramatically different from driving events on an empty country road. They constitute two different environments that support two contrasting driving realities. The second concerns the agent ontology; it concerns what the different agents bring to the scenario in intelligence, function, actions, risk, and roles. Together both the objective and agent ontology of driving determine the decision capacity of the agent. Knowing agent decision capacity is integral to all aspects of informed conceptual analysis. This can only be achieved by investigating agent driving intelligence and how this informs the decision-making capacity, object identification, object value identification, and object value classification. The knowledge or understanding that informs the driving intelligence will be important in determining how the agent perceives the environment and the import of the objects populating it. For example, the driving intelligence of a driver who has only driving knowledge of country roads will influence the ontology of an urban rush hour accident scenario differently from a driver with years of inner city driving experience.

Driving intelligence has an important defining input to the ontology of driving, driving decisions and precollision decisions. The diversity of driving environments, the array of populating actors and artefacts, and the range of the intelligences as the main driving agents determines numerous components that present a complex and diverse array of inputs to the ontology of driving and pre-collision scenarios. The analysis focuses on the question of decision intelligence and investigates how ADI, when contrasted with HDI, determines a different ontology to a driving event in relation to its unique intelligence. The claim that the underlying categorical differences of the two driving intelligences as two paradigms is reinforced when we approach both ADI and HDI confronted with URTA's, as presenting two distinct ontological maps. The first ontological difference concerns the pre-crash window of accident analysis and the second concerns the intentional decision-making that follows. Collectively they frame part of ADI's different and conceptually unique ontology.[19] ADI with the advantage of pre-crash algorithms, collision avoidance, safety and damage limitation technology, will have a window wherein the system can calculate and make decisions. This is ontologically different to the ontology of HDI. HDI rarely has a window of calculation. Rather, the ontology of HDI is defined by an instinctual, stressful, and emotive human response that defines HDI in URTA's (Hermitte 2012, Trappl 2016) and what Sunstein (2005) describes as moral heuristics. Moral decisions that "are moral shortcuts, or rules of thumb, that lead to mistaken and even absurd moral judgments" (Sunstein 2005). HDI is largely blameless in such scenarios due to the overwhelming stress of the scenario. The ontology of ADI consists of predetermined calculated decisions that will determine ADI decisions to be accountable. Without cognisance of these differences, moral frameworks of analysis are under threat of misapplication.

Human Driving Intelligence

HDI is human intelligence applied to the task or context of driving. Human intelligence consists of numerous aspects and behavioural drives. Driving intelligence is defined by our ability to circumvent the numerous driving events we encounter every day. The driving environment is diverse and full of variable values we must effectively assess, judge and make equally diverse decisions in response to the events (Figure 1).[20] Statistically, HDI is remarkably successful at making good driving decisions in responding to the array of driving events encountered in our daily driving workload.[21]

As the above chart elucidates, driving is an extension of our daily human phenomenon. HDI is therefore immensely proficient at surveying, assessing and making decisions in response to environmental variables. Some benefit the function of driving and some undermine it. For example, human reason and problem solving are immensely beneficial to the driving task, while the human behavioural instinct to irrationally respond to an unexpected event brings significant risk to the function of driving. The ability to perceive and identify actors, objects, artefacts, relationships, and risks in part define general human intelligence. The function of driving is an amplification of our normal active interaction with our environment. Driving intelligence is the ability to manoeuvre through the road network while abiding by safety, legal and even ethical determinations. However, it is the introduction of objects and especially objects of value that increases the risk metrics of driving exponentially (Goodall 2014a). This complex engagement of bringing intelligent thought processes to the driving phase contrasts against the predominately-instinctive decisions that

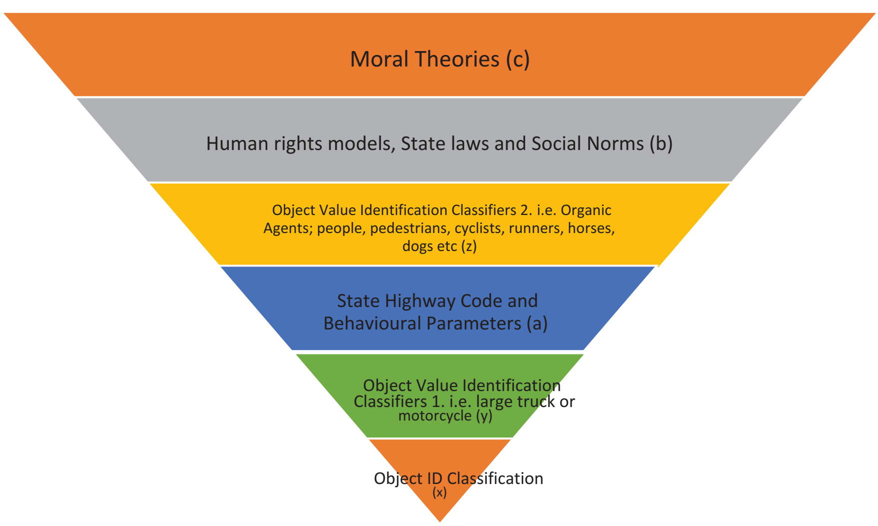

Figure 1. The chart represents a bottom-up break down of the spectrum of intelligence in relation to decision matrixes.

define human driving decisions in response to URTA's. Driving decisions quickly become moral decisions when objects of value and people are introduced into the environment. Risk assessments now contribute to the decisions relating to objects of value. Decisions that involve other parties, the law, and ethical considerations are very much embedded in our driving intelligence. A key challenge to understanding the significance that objects of value play in our driving decisions is difficult to ascertain when one considers the evident relativism involved in human moral intelligence and how different objects are valued according to numerous value systems.

There are at least five differences, which frame the ontology of HDI as different from the ontology of ADI.

(1) Human intelligence is intrinsic to the normal functions of perceptible object interaction and safe and strategic manoeuvrability through the environment.
(2) Human perception concerning object and environmental appraisal is integral to human interaction. This concerns all organic, artefactual, environmental object recognition, value assessment, risk assessment and object-relational understanding are intrinsic capacities to human intelligence.
(3) HDI is a functional application of human intelligence. Driving is a focused application and an applied extension of the inbuilt human capacity to manoeuvre and interact with its environment.
(4) HDI presents a potential application of all human intellectual, behavioural, emotional, and moral forms of intelligence (See Figure 2).
(5) The application of human intelligence to a task such as driving does not define human intelligence rather human intelligence defines the application. For example, in Figure 3 we can remove HDI and driving codes and replace them with other examples of transportation, such as horse riding codes and horse riding intelligence. This identifies the functionality of human intelligence to control tasks such as driving and riding. These tasks represent extensions and focused applications of human intelligence.

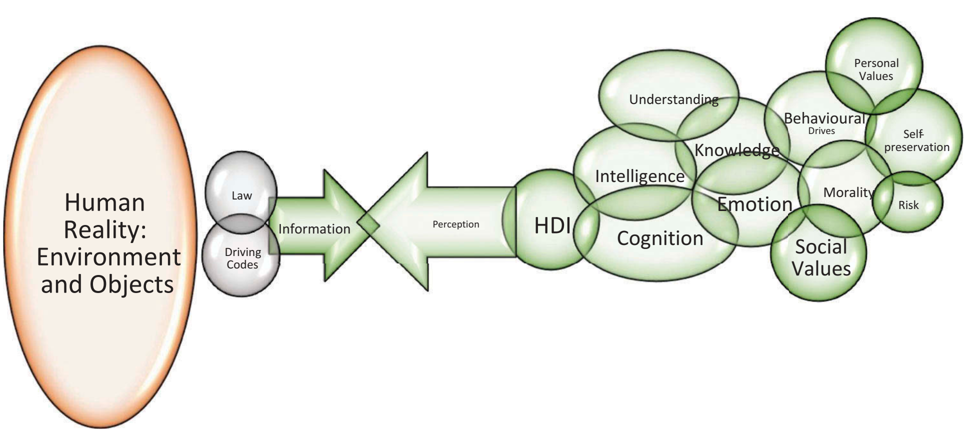

Figure 2. The dynamic and fluid conception of HDI as an application of human intelligence.

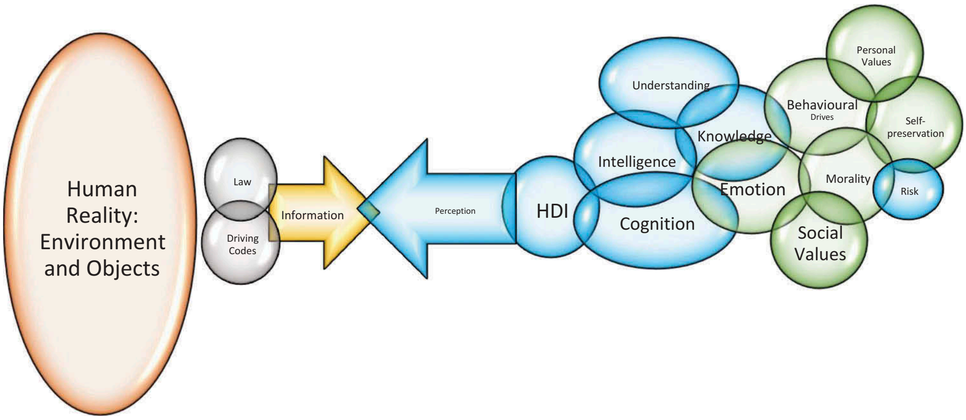

Figure 3. Driving Decisions concerning manoeuvrability will involve perception focused abilities (highlighted in blue), environmental and object intelligence, knowledge and ability. Such decisions rarely require the emotion or moral aspects to human intelligence.

Artificial Driving Intelligence and the Ontology of Unavoidable Road Traffic Accidents

Our human road network presents a fluid diverse environment. Vehicles manoeuvring through this environment will undoubtedly be confronted with unexpected events, potential collisions and life or death RTA's. How ADI responds to such moral scenarios depends upon the intelligence framework that determines its actions. If it has no moral intelligence, then it will respond to moral scenarios as non-moral scenarios and the value spectrum is solely concerned with relational quantifications between the individuals and objects. The ADI will merely respond and make decisions based upon valueless data such as object metrics of size, mass and speed. This could be a negative outcome and could pose a significant threat to using the technology. This concern is developed further by Lin who offers a trolley dilemma as a moral thought experiment. Confronted with a URTA the vehicle will hit and kill both an 80-yr. old lady and an eight-yr. old girl (decision A) unless it swerves to left and just kills the 80-yr. old lady (decision B) or swerves to the right and kills the 8-yr. old girl (decision C). No action means two deaths and action A and B both offer one-death options. Therefore, decision B and C are better, but which one is the right decision and if it is wrong to discriminate between people, choosing B and C is wrong and choosing A is wrong. The moral experiment presents a dilemma and challenge for ADI. How can an ADI respond to this dilemma?

> *"This is a dilemma that is not easily solvable and therefore points to a need for ethics in developing autonomous cars."* (Lin 2015)

The technology for the first time presents a technological window of opportunity. This window presents a technological envelope that captures a momentary snapshot of the pre-collision scenario, wherein possible routes and responses can be assessed. This technological envelope can also be scrutinised post-accident to ascertain accountability and responsibility. Scenarios of unavoidable RTA's bring the question of data into more focus. As the sensor and analytics technology continues to develop, accurate data concerning the people populating the maps will reach a point where individuals are not just identifiable as stationary or moving objects, rather they are quantifiable in numbers, age, and perhaps profile significance to the environment such as a police officer directing traffic. At this level of mapping, the capacity to identify agents in the environment potentially raise the possibility of the ADI being confronted by moral decisions. As more data informs the topographical map, the technological envelope of driving and mapping the terrain, populating the map and calculating trajectories and responding to driving events are building up to the need for more than driving intelligence alone. The level of environment mapping we argue will undoubtedly require contextual analysis from the standpoint of moral intelligence. We claim the availability of more precise data relating to the environmental map will progress the need for moral intelligence. If the map is populated with objects identified as people, then a value weighting is placed upon the objects more so than other organic objects such as dogs or cats. How the driving AI responds or is programmed to respond to these value metrics will determine the moral decisions it makes. ADI ability to identify objects and place values on them will drive the need for moral intelligence.

Therefore, data, the availability of it and the systems to analyse it, will necessitate a need for moral intelligence to be a part of the systems driving intelligence and technological envelope.[22] This goes both ways in the sense that for a machine intelligence not to discriminate and make informed decisions in relation to URTA's it requires information from the topographical map to ensure justice and fairness and to ensure that the decisions must respect all social and legal obligations, and this includes conceptions of equality (Figure 4). This point is echoed by Lin (2015, 71) who asserts that in both German and U.S legislation, there is an emphasis on the right to life and human dignity and an emphasis on equality between citizens. Therefore, to avoid discrimination, ADI also requires sufficient data input to support an informed analysis. The ontology of ADI's capacity to retrieve and process data contrasted against HDI's capacity to make moral, immoral, and instinctive decisions identifies the important differences between them. It identifies how the symbiotic relationship between increasing data inputs combined with improved data analysis will determine a scenario where the topographical map is populated not just with objects but with value spectrums relating to the objects. It is also not only values that may need to be added to the data map but also potential risk and threats too.

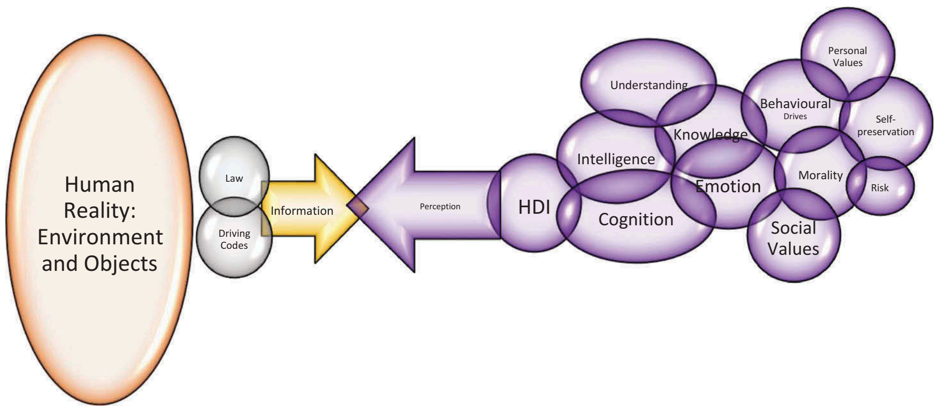

Figure 4. Driving decisions concerning reasoned moral decisions could represent a complex processing of the decision that involves all capacities to assess the decision scenario. It may involve risk combinations, social values, personal values, and emotion. It will undoubtedly involve conflicts between different aspects of the decision-making intelligence. For example, risk may conflict with self-preservation and personal values against the social values and moral values not to harm a child against self-preservation.

Therefore, the topographical map immediately becomes data driven with many layers of data pertaining to object identification, object relations, risks, and threats. This is echoed by Goodall (2016) and Trappl:

> *"Much of the discussion of automated vehicle ethics has focused on hypothetical dilemmas such as the trolley problem and its variants. However, they are not the only problems with ethical implications. Vehicles must constantly assess the value of dangerous actions, and especially in crashes, must compare the values of different objects on the road."* (Trappl 2016)

The topographical representation or rather the reality that ADI perceives is determined by its capacity to gain data from the environment or a perceptible driving event (Figure 5). It is also determined by the capacity to process the acquired data into valuable insights relating to its function and driving telos of safe navigation. Although both HDI and ADI have a great deal in common, such as the shared driving telos of reaching ones' destination safely, the differences outweigh the similarities. The first important difference is the conception of driving intelligence itself. This is evident when one considers the intelligence decision-making capacity of HDI and ADI. The analysis investigates the differences by elucidating the form of the ontological differences that HDI and ADI present. The differences are intrinsic to understanding the concepts of HDI and ADI. The most obvious difference concerns the information and data that HDI and ADI have access to. HDI and ADI are also different in how they process the information and data they receive. Intelligence and the topographical map that forms the core to ADI reality are ontologically different from the visual experience of HDI's reality.

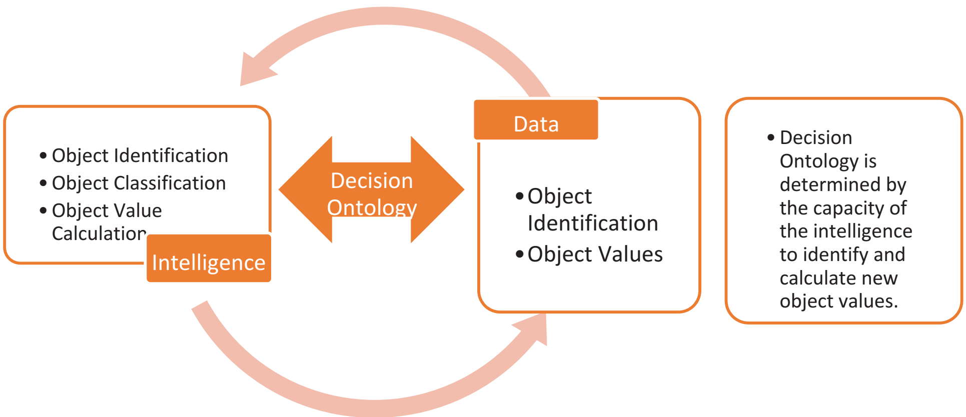

Figure 5. ADI decision ontology.

Mapping the Ontology of a Moral Driving Decision

This capacity of ADI to identify and make operational calculations, according to object values, that it determines has, we claim, significant impact on the ontology of the scenario. This emphasis on the identification of object values and the significance this has on the ontology has also been forwarded by Dogan et al in the AVEthics project (Dogan et al. 2016). Dogan et al, emphasise the importance of the ability to label objects as a process that contributes to the moral ontology of a scenario.

> *"…we will argue that 1) an artificial ethics requires representation and categorization of the morally relevant entities in order to define its "ontology", which is a moral issue per se, 2) an awareness of different entities in the traffic environment could be implemented by assigning to each a numerical value that would be taken into account by the AV control algorithms, and 3) a special "self" value would be added, for an AV carrying humans may not share an ethics of self- sacrifice."* (Dogan et al. 2016)

The format of how HDI has object awareness and places values on objects in its environment is different to how ADI will locate, identify, track, and calculate value metrics to objects. The differences have important ontological determinations to them that will be integral in understanding the technology, its relation to human decision-making, its limitations and capacities. This concern regarding ADI decision-making and object value metrics defines the technology. Especially, in relation to the decision-making intelligence, pre-collision algorithms and how the ADI will respond to unavoidable road traffic accidents. An intelligence confronted with an environment with value objects must make decisions relating to the spectrum of objects, and potential harm, or loss of life. Therefore, the question is, does ADI and its ability to distinguish between objects in a hierarchical format necessitate moral decision-making.

Such decisions have been addressed in recent literature in the context of the Trolley Dilemma (trolley dilemma).

HDI is defined by its ability to make decisions in response to driving events; these decisions consist of a broad spectrum of varying ability from the instinctual to rationally informed decisions. Throughout the HDI and the spectrum of decisions, moral decisions present our moral intuitions and considered responses to driving events. Robert Trappl, a computer scientist with a background that focused on the ethical context of AI, posits an insightful account that highlights the reality of URTA's and the need to consider trolley dilemma's in the context of HDI (Trappl 2016). Trappl makes an important point that identifies a problem space that remains needing further investigation. It concerns the area of URTA and ADI moral decision-making capabilities by further investigating the difference between HDI and ADI decisions. The difference centres on the ontology of HDI's capacity to respond to URTA's and ADI's potential capacity to respond to URTA's.

> *"In the winter of 2013, a report about an accident was published in most Austrian newspapers: The driver of a school bus saw a deer jump from a nearby forest onto the road in front of the bus. In order not to hit the animal, the driver turned his bus sharply. The bus skidded off the snowy road, rolled down a steep meadow until trees stopped it. Several of the school children were severely injured and had to be flown to hospitals. It was considered a great luck that no child died. The report did not inform about the fate of the deer."* (Trappl 2016)

Trappl argues that the driver's decision not to hit and harm the deer was an ethical one possibly forged many years before the accident (Trappl 2016). This is essentially a human judgement call and represents the distinctly human ability to "exercise judgement in a wide range of dynamic situations" (Lin 2013). Trappl claims the driver's reaction to avoid the deer and swerve the bus off the road results from his own inherent ethical system not to harm animals (Trappl 2016). Unfortunately, this reaction which reflected his ethical view of not harming animals caused great harm to himself and his schoolchildren passengers, who he had to care for. Both rupture and emergency pre-collision phases are generally identified as unexpected driving events that go beyond the skill and proficiency of HDI. The inability of HDI to react to unexpected events and stressful emotive scenarios means that instinctive human reactions are not to blame for making a wrong decision. Human drivers often make decisions that are nothing more than instinctual reactions to situations; more often than not, these instinctual reactions have adverse consequences to them. However, ADI does not have this means of avoiding responsibility for a decision it makes. Rather, the scenario is different for ADI given it will not only have a window of opportunity, but its pre-collision decisions will be transparent and open to interrogation. This point is echoed by Lin (2013, 2016) and Trappl:

> *"An important difference between the decision of a human driver and a self-driving car is this: A human driver may be excused or be forgiven for making a wrong decision, i.e., a decision considered unethical in retrospect because of the stressful situation in which s/he had to decide. An intelligent software program has no such excuse." (*Trappl 2016)

Trappl identifies the important difference between human instinctual responses, which is often legally accepted as characteristically human instinctual and blameless response (Trappl 2016). This view is supported by modelling in accidentology, which also claims that confronting a URTA, is a stress response that denotes an overloading of information beyond the ability and intelligence of the driver (Hermitte 2012, Van Den Beukel and Van Der Voort 2016). HDI responses to URTA's are typically irrational, emotive, instinctive and often predetermined by what Trappl suggests to be established ethical systems of moral values. The ontology of the scenario is framed by the HDI and the capacity to respond to the URTA. Although other decisions were possible for the bus driver, such as hitting the deer or swerving in a different direction, these decisions were not psychologically available to him. His reaction to the driving event was determined by a deep-rooted moral belief system not to harm animals (Trapp: 2016).

Many investigations that have centred upon utilising the trolley dilemma have failed to consider the compatibility of the experiment specifically in relation to how intelligence determines ontology (Goodall 2014a, Lin 2013, 2015, Millar 2014, 2016, Nyholm and Smids 2016). They have taken a conceptual experiment from the context of human intelligence and decision-making and have attempted to apply it to a context of artificial intelligence and decision-making. This is evident when one considers that human intelligence is inseparable from moral intelligence, whereas artificial intelligence and ADI are devoid of any moral intelligence. The identification of these differences will support a more accurate assessment of the moral context to ADI confronted with URTA's. Examining this process allows one to make a meaningful ontological distinction between HDI and ADI. It also allows us to better understand the juxtaposition of various inputs and posit a more informed role for ADI and the possibility of artificial moral intelligence.[23] This relationship between intelligence and moral decisions is inherent to the challenge of understanding ADI and URTA's.

The ontology is determined by the ability of the intelligence to not only, identify value objects but also to calculate, and autonomously prescribe value metrics to objects. This forensic ability to identify values and calculate object values, enables hierarchical calculations regarding which objects are to be avoided and which objects can be considered hard and soft targets.[24] This intentional capacity to make informed object identifications in relation to values will also determine ADI to be responsible for its object value identifications[25] and will be accountable for its object targeting decisions.[26] Quantifying scene intelligence, therefore, concerns the ability of ADI to determine object values. The capacity of both HDI and ADI to identify object values, calculate object relations in terms of new values, and quantify risk metrics in relation to decision planning, we claim will determine the ontology of a scenario. What HDI and ADI bring to the scenario as an intelligence that processes object values determines a unique intelligence specific ontology. The actor as an artificial intelligence has a unique determination on the ontology of a scenario.

Conclusion

Human intelligence is not only intrinsic to our moral decisions but when we frame our moral decisions we are also framing human intelligence and all that comes with it. To utilise the trolley dilemma to analyse AI and (the potential of) artificial moral intelligence without awareness of the ontological frame, anthropomorphism and human bias that the concept of AI brings to understanding machine decisions in the context of driving, will only lead to misapplication and misunderstanding. The threat of misapplication relates to the categorical difference between human contexts of meaning and moral analysis. However, for the moment, the important contribution that the paper offers concerns how the technology can be further developed with the awareness that human moral experiments are not ontologically compatible with artificial intelligence. The URTA challenge presents scenarios to ADI that will sometimes require complex analysis, decision responses and assessment of possible human harm and fatalities. In developing the analysis, the paper has focused on defending the claim that before any ethical analysis can take place there is a requirement to utilise the trolley dilemma as an ontological tool. The identification of the specific ontology of the scenario is necessary for elucidating concepts of liability, risk, accountability, and responsibility. In ADI and machine moral intelligence, these concepts have different applications to HDI. Ontological elucidation, through the prism of the trolley dilemma is the first step in anticipating the ethical challenges that the emerging technology of ADI present.

As autonomous vehicles edge nearer to becoming a part of our chaotic human road networks all actors need to better understand the decision capacity of ADI. The hope is that the technologies and the decision capacity will bring order, control and safety to the road networks. The technologies are claimed to be ethically justifiable in this respect. However, this does not remove or lessen the need to seek better understanding of machine decisions. Especially, decisions that concern human life and welfare. There is a need to understand the decision-making spectrum that defines the intelligence and algorithms of ADI to identify its abilities and limitations. This is based on two inferences relating to the ontology of URTA's. Each inference amplifies the core challenge of HDI as a challenge of moral decision intelligence that confronts ADI. First, concerns the nature of the road network itself as an obvious source of driving events that will demand moral intelligence. Both HDI and ADI share two important characteristics; they will share the same diverse road network and will be governed by the same rules, ethics and practices supporting safe road use. The road network will also present the same diverse volume of driving events to ADI as HDI has confronted. Second, the ADI will be accountable for the decisions it carries out in response to URTA's and as the technology develops more and more environmental data will be delivered to the ADI. This data will include data on the objects and agents populating the environment. This data will be necessary for improved decision-making and will undoubtedly involve data on the human agents that will confront the ADI with what we would identify as moral decisions. For example, when confronted with a URTA should the ADI avoid the group of children and target the motorcyclist as the lesser of the two potential collisions? Data in relation to human agents populating the environment will confront the ADI with moral decisions. The ability of the ADI will determine how it intelligently uses the data available. Clearly sometimes appealing to an ethical system can constitute the wrong output, as Trappl's example communicates. This is primarily because of the individual relativism in relation to what values we consider important and hold more highly than others. One of the important outcomes of the research is in positing a coherent account of HDI as consisting of interwoven capacities of driving and moral intelligence. To respond to the diverse human road network, it would be advantageous to have a sufficiently advanced moral intelligence that could effectively respond to URTA's. Moral machine intelligence is necessary to respond to the array of variable road events that will undoubtedly confront the ADI controlling the vehicle. Benefits of moral machine intelligence, as a part of ADI, relate to; negating instinctual human responses to URTA's, offering a more informed and accurate response to URTA's by building upon the benefits of ADI precise operation and environmental mapping, and potentially reducing individual human biases that impact upon the efficiency of pre-crash decision making.

Notes

1. The term automated vehicle technologies (AVTs) is used to refer to the array of technologies that are currently used in automating driving workloads. Following Dogan et al, it is evident that the combination of technological envelops increase automated driving, by reducing the period of the human driving required to be kept in the driving loop (Dogan et al. 2016).

2. ADI is an important classification because it provides some insulation from current controversies relating to AI. It provides a context of meaning and elucidation specific to the technological application. Although, for the present purpose the categorization is largely heuristic.
3. In his paper "Moral Competence in Robots?). Malle claims that moral competence is more than moral agency and the capacity to determine and act according to what is judged to be right or wrong. Competence is according to Malle a more beneficial ascription as it supports a meaning that highlights the ability to carry out a task as an aptitude, qualification or capacity. I follow Malle in developing this point further by making the distinction stronger and specifying it as moral intelligence (Malle 2014).
4. There is also the concern that informed decision making in response to URTA is equivalent to targeting (Lin 2015).
5. The question of ADI and capacity to respond to complex driving events and decisions that define HDI's daily journeys as well as the possible array of different URTA's also leads to concerns regarding accountability and liability.
6. Public trust and risk perception is currently focused on the improved safety benefits that AVTs potentially offer. However, accurate risk perception is being undermined by inaccurate portrayals of trolley dilemma's as definitive examples of safety hazards relating to AVTs. See: Young (2016). "The Moral Algorithm: How to Set the Moral Compass for Autonomous Vehicles Moral Decisions by Autonomous Vehicles and the Need for Regulation". Available at: https://gowlingwlg.com/getmedia/0eb5a71b-37fb-4ea9-a2c5065fbc0d1e10/161205-the_moral_algorithm [Accessed May 9, 2017].
7. Accurate risk metrics are also necessary for ethical analysis, given that risk and ethical analysis collectively contribute to the development of frameworks of governance.
8. Cheng (2011) suggests that we can understand the goal-focused system as one designed to reach waypoints and the final destination in a safe manner.
9. John McCarthy MIT 1956 see: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html and http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf .
10. Selman et al. (1996) identify numerous challenges that face the development of AI and many of the challenges relate to the inability of AI especially as automated robotics in human environments to adapt to the environments. Organic intelligence is excellent at adapting if it does not adequately adapt it dies.
11. There is a historical challenge to modern ethics coming from Elizabeth Anscombe (1958) and Richard Joyce (2003) that maintains that most if not all ethical analysis is deeply confused. This is evident in its attempt to rationalise and systematise emotive scenarios.
12. Risk assessment is a precursor to informed ethical analysis. Ethical analysis cannot move from a phase of anticipatory hypothetical analysis to analysis of evident ethical issues until the risk assessment has furnished data to identify the complex web of relations relating to what risks relate to the use of the technology. For example, there is a clear risk of user error relating to automated vehicle technologies and informed decision making regarding the capabilities and limitations of the technology. This point was raised in reference to Tesla Autopilot system software as a deceptive name that suggests like an aircraft the system can take over full automated control.
13. Both private and public policy governing effective regulation and support for the development and deployment of an emerging technology is dependent upon adequate risk and ethical data to contribute to a framework of governance to support the significant scope in supporting legislation. (for more on the intricate relationship between risk, ethics and policy see: Asveld and Roeser 2009; Anderson and Nidhi et al. 2014, Chopra and White 2011, Kumfer et al. 2016) .

14. There are three variations in development concerning vision based autonomous driving systems (Chen et al. 2015). It is therefore likely that developing autonomous vehicle technology will depend upon visual data analysis. All vision based system functionality is based on the system's ability to retrieve external visual data it can analyse to formulate decision options on trajectories and assess possible risk metrics within the specific technological envelop of each action. Vision systems rely on effective data acquisition to formulate informed and accurate routes. The paper is part of a European Horizon 20/20 research program focusing of the development of a Vision Inspired Driver Assistance System (VI-DAS).
15. Like human intelligence, ADI relies upon the capacity to identify objects, appraise object values, calculate, and carry out control outputs as driving decisions. It is widely recognised that ADI will act as an autonomous agent whose actions will be indistinguishable from many of the actions of human driving.
16. This is evident in the case of Artificial Intelligence (AI), which has proved problematic since John McCarthy first coined the terms in the 1950's. McCarthy was aware of the dangers of the use but nonetheless forwarded it by emphasising that "For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving." (McCarthy et al. 1955). http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html .
17. Cited in (Goodall 2014a): Fraichard, T., and H. Asama. Inevitable Collision States: A Step Towards Safer Robots? *Advanced Robotics*, Vol. 18, No. 10, 2004, pp. 1001–1024.
18. In road accident and behaviour research it is accepted that "traffic accident involvement is more closely related to human judgement and decision-making than the mere inability to control the vehicle, and therefore, the focus of driver behaviour and decision-making patterns became a popular research area in road-safety applications" (Meiring and Myburgh 2015).
19. Anat Biletzki claims that concepts can be pushed beyond the limits of established meanings when technological innovation forces change in how we perceive the world and determines different social interactions. I claim that this approach is similar to the conceptual frame that ADI requires to further analysis. (Biletzki 2013) .
20. "It is possible to drive a private car 13000km a year for 50 years with more than a 99 % chance of survival" (van Suntum 1984: 160). For a more recent breakdown of RTA figures and related causes, see figures at https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115.
21. For example appealing to German RTA figures between 2005 and 2009, Hevelke and Rumelin claim there was one accident every 1.46 million kilometres travelled (Hevelke and RumeLin 2015).
22. See (*https://ai100.stanford.edu/2016-report/preface*) .
23. Immanuel Kant in his famous work "Groundwork for Metaphysics of Morals" emphasised the important relation between the human faculty of reason and the fundamental "ought" that relates to it, the capacity to reason, intelligence and morality are interwoven (Kant 1783).
24. https://en.wikipedia.org/wiki/Soft_target.
25. Brooks et al. (1996) identifies one of the challenges that confronts AI concerns the ability "To be successful in realistic environments, reasoning systems must identify and implement effective actions in the face of inescapable incompleteness in their knowledge about the world. AI investigators have long realized the crucial role that methods for handling incompleteness and uncertainty must play in intelligence." (Selman et al. 1996) .

26. Lin claims that Crash optimisation means 'targeting' and he maintains that the AVTs intelligent system making decisions relating to choosing one object over another is targeting (Lin 2015, 73). We claim that targeting will be a necessary feature of the ADI capacity to make decisions in response to URTA's. The challenge is to understand the intelligence that constitutes the targeting decisions.

Funding

This work was supported by the Horizon 2020 [690772].

References

Anderson, James M., Nidhi K., Stanley, K. D., Sorensen, P., Samaras, C., and Oluwatola, O. A. 2014. *Autonomous vehicle technology: A guide for policymakers*. Santa Monica, CA: Rand Corporation.

Anscombe, G. E. M. 1958. Modern moral philosophy. *Philosophy* 33:1–19. doi:10.1017/S0031819100037943.

Asveld, L., and S. Roeser. 2009. *The ethics of technological risk*. London: Routledge.

Bentham, J. 1843. Rationale of Reward, Book 3, Chapter 1. In *The Works of Jeremy Bentham*, ed. by J. Bowring, vol. 8. Edinburgh: William Tait.

Biletzki, A. 2013. *Meaning as use in the digital turn*. Accessed March 4, 2017. http://wittgensteinrepository.org/agora-ontos/article/view/1899.

Bonnefon, J. F., A. Shariff, and I. Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352 (6293):1573–76. doi:10.1126/science.aaf2654.

Bonnefon, J.F., Shariff, A., and Rahwan, I. 2015. *Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars?* Accessed October 28, 2015. http://arxiv.org/pdf/1510.03346v1.pdf.

Brooks, R., Selman, B., Dean, T., Horvitz, E., Mitchell, T. M., and Nilsson, N. J.(1996). 'Challenge problems for artificial intelligence (panel statements)', in *Proceedings of AAAI-96*, pp 1340–45. Massachusetts, AAAI: MIT Press.

Chen, C., Seff, A., Kornhauser, A., and Xiao, J. 2015. 'Deep Driving: Learning Affordance for Direct Perception in Autonomous Driving.' *IEEE International Conference on Computer Vision (ICCV)* 2015: 2722–30. doi:10.1109/ICCV.2015.312

Cheng, H. 2011. *Autonomous intelligent vehicles theory, algorithms and implementation.* London: Springer.

Chopra, S., and L. F. White, (2011). '*A legal theory for autonomous artificial agents*' Accessed March 12, 2017. https://books.google.com/books?hl=en&id=J5IoSzOKZlUC&pgis=1.

Coeckelbergh, M. 2016. Responsibility and the moral phenomenology of using self-driving cars *Applied Artificial Intelligence* 30 (8):748–57. doi:10.1080/08839514.2016.1229759.

Dogan, E., Chatila, R., Chauvier, S., Evans, K., Hadjixenophontos, P., & Perrin, J. (2016). Ethics in the design of automated vehicles: The AVEthics project. Accessed March 21, 2017. http://ceur-ws.org/Vol-1668/paper2.pdf.

Foot, P. 1978. The Problem of Abortion and the Doctrine of the Double Effect. In *Virtues and Vices and Other Essays in Moral Philosophy*, ed. P. Foot, 19–32. Berkeley: University of California Press.

Gerdes, J., and S. Thornton. 2015. Implementable ethics for autonomous vehicles. In *Autonomes Fahren*, ed. Markus Maurer, J., Lenz, B., & Winner, H., 87–102. Springer

Vieweg, Berlin, Heidelberg. Accessed February 8th, 2017. http://link.springer.com/10.1007/978-3-662-458549_5_OnlinePDF.

Goodall, N. 2014a. Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board* 2424:58–65. doi:10.3141/2424-07.

Goodall, N. J. 2016. Away from trolley problems and toward risk management'. *Applied Artificial Intelligence* 30 (8):810–21. doi:10.1080/08839514.2016.1229922.

Hermitte, T. (2012) *Review of accident causation models used in road accident research of the EC FP7 project DaCoTA*. Accessed January 15th, 2017. http://www.dacota-project.eu/Deliverables/DaCoTA_WP5_D5_9_Review_of_Accident_Causation_models_vf.pdf..

Hevelke, A., and J. Nida-Rümelin. 2015. Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics* 21 (3):619–30. doi:10.1007/s11948-014-9565-5.

Joyce, R. 2003. *The myth of morality*. Cambridge: Cambridge University Press.

Kant, I. 1783. 1985. *Groundwork to the metaphysics of morals, trans*, H. J. Paton. Hutchinson: London.

Kumfer, W. J., S. J. Levulis, M. D. Olson, and R. A. Burgess. 2016. A human factors perspective on ethical concerns of vehicle automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 60:1844–48. Accessed February 1, 2017. http://journals.sagepub.com/doi/pdf/10.1177/1541931213601421.

Lin, P. (2013). *The ethics of saving lives with autonomous cars are far murkier than you think*. Accessed January 15, 2017. http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars.

Lin, P. 2015. Why ethics matters for autonomous cars. In *Autonomes Fahren*, ed. Maurer, M, Gerdes, J. C., Lenz, B., and Winner, H., 69–85. Germany: Springer.

Malle, B. 2014. Moral competence in Robots. *Frontiers in Artificial Intelligence and Applications* 273:189–98.

Maurer, M., Gerdes, J. C., Lenz, B., and Winner, H. 2016. *Autonomous driving: Technical, legal and social aspects*. Berlin: Springer Open.

McCarthy, J. (1996) *What has AI in common with philosophy?* Accessed January 9th, 2017. http://wwwformal.stanford.edu/jmc/aiphil.pdf.

McCarthy, J., M. L. Minsky, N. Rochester, and C. E Shannon. 1955. A proposal for the dartmouth summer research project on artificial intelligence. *Ai Magazine* 27 (4):12–14.

Meiring, G. A., and H. C. Myburgh. 2015. A Review of Intelligent Driving Style Analysis Systems and Related Artificial Intelligence Algorithms. *Sensors (Basel)* 15(12): 30653–30682. doi:10.3390/s151229822

Millar, J. 2014. 'Proxy prudence: rethinking models of responsibility for semi-Autonomous robots. *Proceedings of We Robot*. Coral Gables, FL: The University of Miami School of Law. Accessed April 7th, 2017. http://robots.law.miami.edu/2014/wpcontent/uploads/2013/06/Proxy-Prudence-RethinkingModels__-of-Responsibility-for-Semiautonomous-Robots-Millar.pdf.

Millar, J. 2016. *The momentous advance in artificial intelligence demands a new set of ethics*. Accessed May 11th, 2017. https://www.theguardian.com/commentisfree/2016/mar/13/artificial-intelligence-robots-ethicshuman-control.

Nyholm, S., and J. Smids. 2016. 'The ethics of accident-algorithms for self-driving cars: An applied trolley problem?'. *Ethical Theory and Moral Practice* 19 (5):1275–89. doi:10.1007/s10677-016-9745-2.

Rose, B. (2016). *The myth of autonomous vehicles*.(Accessed March 15th, 2017) https://techcrunch.com/tag/ethics/.

Selman, B. Brooks, R. A., Dean, T., Horvitz, E., Mitchell, T. M., & Nilsson, N. J. (1996). Challenge problems for artificial intelligence. *AAAI'96 Proceedings of the thirteenth national conference on Artificial intelligence*, Portland, OR, Volume 2 pp 1340–45.

Sunstein, C. 2005. Moral heuristics (2005). *Behavioral and Brain Sciences* 28:531–73. doi:10.1017/S0140525X05000099.

Thomson, J. 1985. The trolley problem. *The Yale Law Journal* 94 (6):1395–415. doi:10.2307/796133.

Trappl, R. (2016) Ethical systems for self-driving cars: An introduction. *Applied Artificial Intelligence* 30 (8):745747.

Van Den Beukel, A. P., and M. C. Van Der Voort (2016) 'Driving automation & changed driver's task - Effect of driver-interfaces on intervention'. *IEEE Intelligent Vehicles Symposium, Proceedings* (IV) 2016, Gothenburg, Sweden.

Van Suntum, U. 1984. Methodische probleme der volkswirtschaftlichen bewertung von verkehrsunfaellen. *Z Verkehrswiss* 55 (3).

Walker-Smith, B. (2012). http://cyberlaw.stanford.edu/blog/2012/03/driving-perfection

Wallach, W., and C. Allen. 2008. *Moral machines: Teaching robots right from wrong*. New York: Oxford University Press.

Wang, J., and X. Wang. 2011. An ontology-based traffic accident risk mapping framework. In *Advances in spatial and temporal databases. SSTROLLEY DILEMMA 2011. Lecture notes in computer science*, ed. Pfoser D., Tao, Y., Mouratidis, K., Nascimento, M. A., Mokbel, M., Shekhar, S., and Huang, Y. A. N. vol. 6849, 21–38. Berlin, Heidelberg: Springer.

Young, S. (2016). *The moral algorithm: how to set the moral compass for autonomous vehicles moral decisions by autonomous vehicles and the need for regulation*. (Accessed May 9, 2017) https://gowlingwlg.com/getmedia/0eb5a71b-37fb-4ea9-a2c5-065fbc0d1e10/161205the_moral_algorithm